THE IMPACT OF AN AGING WORKFORCE ON SYSTEMS SAFETY AND RELIABILITY AND ADDITIONAL ERROR-PRODUCING CONDITIONS AS APPLIED TO AN EXISTING HUMAN RELIABILITY ASSESSMENT METHOD

J.C. Williams, J.L. Bell, N.D. Warren and C. Daniels
Health and Safety Laboratory

Some process safety assessments can be extremely sensitive to human error and may be affected by an aging workforce. Even though there is some evidence that human reliability can improve as a result of experience, a wide range of systematic studies have shown that personnel involved in performing detection, recall and recognition tasks will be likely to be approximately 20% less reliable for every ten years age and that, within very narrow limits, this impact will be consistent over the range 25 to 85 years. Other Error Producing Conditions, for example, time-of-day, also show similar predictable reliability degradation and the implications of these findings are discussed.

INTRODUCTION

Currently, when performing Quantified Risk Assessments (QRA) in order to determine what the major components of risk might be, it is normal to examine the principal failure cases for their susceptibility to human failures with respect to the initiation, control, consequence, impact and mitigation of foreseeable accident sequences and, where appropriate, quantify the impact of human error. Even though assessors work hard to ensure that their predictions are as accurate as possible, underestimation of human error by as little as a factor of four can, on occasion, change top event frequencies by more than a factor of ten (Brookhaven National Laboratory, 1989; Samanta et al. 1990). Delboy et al. (1991), for example, have shown that process safety can be very sensitive to human error (sometimes by factors of well over a hundred), so even small changes in human reliability can, in some circumstances, have profound implications for plant safety.

Process safety assessments also show that assumptions about human performance can have profound effects on the way in which risks would be managed. The work of Wong et al. (1990), for example, shows that human performance, and, in particular, human variability can change process safety assessments in powerful ways. Slight changes in nominal assessments can result in massive differences in the computed risk values that an overall assessment might otherwise produce. Samanta et al.'s theoretical studies show that if estimates of human error are underestimated, the calculated effects on hazardous releases will be very great, whereas if the estimates are over-estimated the effects will be comparatively small. Clearly, as has been illustrated by Brown et al.

(1990), such differences are non-trivial, and could result in entirely different risk management decisions being taken, simply as a consequence of the computational process. Therefore, having confident estimates of human reliability and the influences on human reliability is clearly an important issue in relation to managing risk. So, any improvements to existing human reliability techniques should prove beneficial and be welcomed.

For 20 years, in the U.K., the HEART (Human Error Assessment and Reduction Technique) (Williams, 1985 and 1988) has been used extensively by the Human Reliability Assessors' (HRA) community to assess and quantify the impact of human unreliability on system reliability. For many, the method provides meaningful insight to detailed technical assessment and reasonably accurate prediction of the likelihood of successful human – machine interaction. In the mid '90s, the work necessary to validate the method was undertaken via independent research (Kirwan, 1997) and the precision achieved was found to be comparable to, but not quite as good as, other forms of systems reliability assessment (Snaith, 1981).

Over the years, a large number of applications have been made and the HRA community has become comfortable with the technique and its underlying theory and principles. One of the most powerful features of HEART is the fact that it provides evidence-based information regarding the probable impact of Error-Producing Conditions (EPCs) on human performance and how these EPCs might combine to threaten human reliability and system safety. However, there are known shortcomings in the method and areas of incompleteness and, when it was first published, these were made apparent to the unwary. Also, as the method has matured, a number of researchers have tried to minimise the consequences of these shortcomings, but, it would seem, with varying degrees of success so far. Therefore, there is a need to reduce the uncertainty in the method's use and underlying features, a need to update it and, where possible, a need to make the method more usable.

Not all the industrially-relevant EPCs in HEART have been identified, as yet, and some of those that have been identified are not supported by as much evidence as one would like to have in order to have confidence in an assessment in order to take appropriate action. In addition, over the last 20 years, the human factors literature has grown enormously, with EPCs that were previously largely-unresearched now becoming amenable to incorporation into the model and with those that, to start with, had limited support now becoming much more easily analysed and assessed.

The net result of all the above is that there are many improvements that can now be made that were simply infeasible in the mid '80s. These include enhancements to content and applicability, as well as to the usability by non-specialists.

HRA DEVELOPMENTS

Developments that are being encouraged in the UK HRA area, at present, include a perceived requirement for enhanced sensitivity of methods to Performance-Shaping Factors. This includes, reduced and extended operator action times, fatigue, working hours, shift patterns, time-of-day, and general work organisation and management factors such as team interaction and the impact of safety cultures. Although there is also some interest in developing customized versions of HRA methods for specific industries, the purpose of this particular study was to extend the general-purpose nature of the HEART method by capturing data for those EPCs that are currently under-substantiated and, where feasible and appropriate, to extend the number of EPCs available to the practitioner.

For this particular study, two EPCs were selected for attention, Aging and Time-of-Day. The first of these is already in the original HEART method as EPC 38. The second, Time-of-Day, was not in the original method, but since the method was first published, information has become available that allows us to be more accurate about the implications of circadian effects for human reliability than hitherto, so this, too, has been quantified.

METHOD

To gather information relevant to Aging, a very wide-ranging literature search was conducted. This included a detailed and painstaking search of a range of specialist journals, including *Psychology and Aging*, *The Journal of Gerentology* and direct contact with the principal researchers of the day. The literature search flagged up some very important developments which had occurred since the mid ’80s.

In total, some 82 relevant papers have been identified to date and of these, some 15 were found to be usable in terms of the data that could be extracted for further processing. Data were regarded as being usable if the results of the various experiments could be plotted as probabilities of failure, if the experimental sample sizes were substantial (above 10 Subjects per study), and if age was a factor in the experimental design.

The range of experimental studies considered was very great and it is noticeable that, over the last 20 years, the information on Aging and the Impact of Aging on Human Performance that was needed to “complete” the HEART model in this respect has now largely become available, in a way, and to an extent that was simply unthinkable in the mid ’80s.

Each experiment identified as being potentially usable was analysed for evidence of the effect of age in relation to detection, recall and recognition. Task failure data were converted to probabilities of failure and these were plotted on a logarithmic-scaled spreadsheet for initial visual comparison. The wide range of studies used quite different measures which, in some cases, produced an error probability but, other results were presented as a percentage of information lost or the percentage below the maximal score. For the purposes of this study, all results were converted to a common measure of human performance between 0 (best) and 1 (worst), which is interpreted as an error probability.

After the first few data sets were plotted it became apparent that some regularities seemed to be present in these particular sets and, accordingly, some further more robust statistical techniques were deployed in order to find out whether these apparent regularities had any substance in fact.

STATISTICAL METHODOLOGY

In order to provide a common basis for analysis across a range of studies, to permit meaningful comparison, data from the experiments of Ardila et al. (2000), Atchley and Kramer (2000), Baddeley et al. (1994), Detweiler et al. (1996), Gamboz et al. (2000), Kirasic et al. (1996), Kvavilashvili et al. (2005), Maylor and Lavie (1998), Ostrosky-Solis et al. (1998), Parkin and Walters (1991), Roge et al. (2003), Salthouse et al. (1996), Salthouse (2004), Sliwinski et al. (2003) and Whiting et al. (1997) were calculated as "odds".

$$(\text{odds} = p/1 - p), \quad \text{where p is the probability of an error occurring.}$$

These "odds" were then plotted as natural logarithms (Fig. 1).

It will be apparent via visual inspection alone that there is a consistent change in odds that is associated with age and that this change is relatively uniform across a wide range of ages. This means that reliable predictions can be made about the impact of age on human reliability.

This analytical approach is consistent with the widespread use of logistic regression to model the probability of an event occurring where dichotomous observations are available. For our analysis, direct observations of the occurrence of errors are unavailable and logistic regression inappropriate. Instead, our analysis uses a weighted linear regression

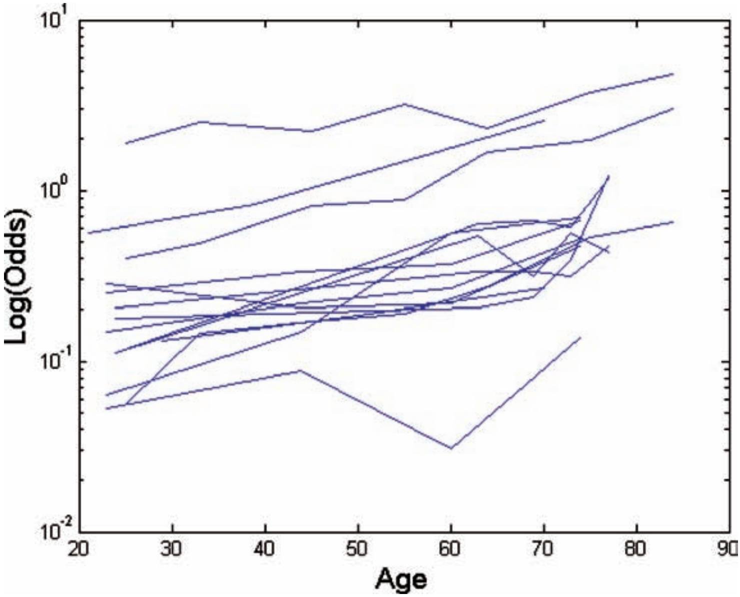

Figure 1. Probability of failure data taken from each relevant study

model to examine the effect of age on the log odds. The weight for each data point is determined by the number of subjects in the underlying experiment with the variance of each data point's error term being inversely proportion to n (because each data point is the mean of n subjects). Using weighted regression ensures that, despite widespread variation in the size of studies, appropriate conclusions are drawn. The model for the log odds can be written as:

$$\log(\text{odds}) = \mu + \text{task effect} + \text{age effect} \times \text{age}$$

Here μ represents some nominal base line value for the log odds over all studies, the task effect describes the adjustment for a specific task and the age effect the incremental effect for every year of age. The age effect determined by this analysis represents an integrated estimate of the effect of age based upon the entire set of experimental studies. A 10 year increase in age increases the odds by a factor of

$$\text{Exp}(10 \times \text{age effect})$$

This approach deals effectively with data which have large values corresponding to probabilities close to one. This is a major advantage for it allows the model to be extended to multiple error producing factors without the danger of their combining to give impossible results. It is worth noting that the probability of an error can be easily recovered from the odds using:

$$p = \frac{\text{odds}}{1 + \text{odds}}$$

For rare events, where the probability p is small, the probability and the odds are almost numerically equivalent.

Work is currently underway to develop a mixed effects model to investigate variation in the effect of age with different generic task groups. This analysis will also model the dependencies between 'data points' taken from the same study and investigate the impact of some additional Error Producing Factors.

SUMMARY OF THE EFFECTS OF AGE ON HUMAN RELIABILITY

The coefficient for age in the current weighted regression model is 0.015 (95% confidence interval 0.011–0.019, p-value <0.0001). This value implies a relative increase in the odds of an error occurring of 1.16 or 16% (95% confidence interval 12%–21%) for every 10 year increase in age. For small error probabilities (of the order 10^{-2}) this corresponds to a 16% increase in the error probability itself.

Finally, some experimental results were kept to one side, so as to determine whether the impacts predicted by the data treatment would actually predict the outcome in the unanalysed experiments (Detweiler et al. 1996; Glisky et al., 2001; Fernandes & Moscovitch,

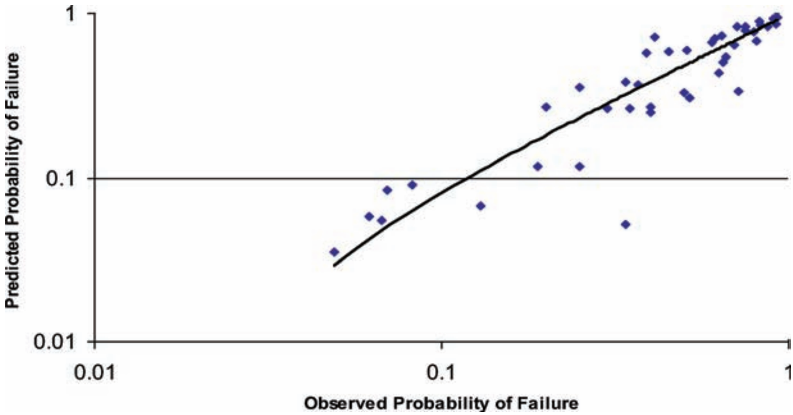

Figure 2. Correlation between observed and predicted experimental results

2003; Kausler & Puckett, 1981; McIntyre & Criak, 1987; Madden & Langley, 2003; Mantyla & Backman, 1992; Maylor et al. 1999; Park et al. 1989; Rankin & Collins, 1985; Shinar et al. 2005; Singer et al. 2003; Taylor et al. 2005). Evidence from this withheld information suggests that some fairly accurate predictions can be made, based solely on the strength and direction of the rest of the data set. These are shown in Fig. 2.

The correlation coefficient is .916, $p < .01$ which suggests that accurate predictions of the impact of age on human reliability can be made based solely on an EPC of 1.2 for every 10 years age, over the range 25 to 85* years (*non demented). Further work to refine this EPC is currently underway, but it is not expected to depart radically from initial indications.

TIME-OF-DAY EFFECTS

A second EPC was also investigated, to a lesser extent, as part of this study. Provisional investigations and assessment of the impact of time-of-day on human reliability suggest that the variation between the diurnal high and low of performance is about a factor of 2.4; the highest reliability being around 16:00 hrs and the lowest reliability, somewhat unsurprisingly, at around 03:00 hrs. The "neutral" point is at around 20:00 hrs. There is some evidence of a "post lunch dip", but, in overall reliability terms, this phenomenon is not very great, by comparison with the overall circadian change in reliability. It contributes about a factor of 1.3 to overall human unreliability specifically at around 14:00–15:00 hrs. The data analysed and consolidated come from a range of sources, including laboratory, field and accident studies by Browne (1949), Bjerner et al. (1955), Blake (1967), Hildebrandt et al. (1974), Harris (1977), Nicholson et al. (1984), Craig and Condon (1985), Borland et al. (1986), Mittler et al (1988), Howarth et al (1989) and Folkard (1997).

With the knowledge that there is a measurable impact on human reliability associated with time of day, it is now possible to examine major hazard faults. For example, the "overfilling a tank" fault identified in the risk assessment by Khan et al. (2001), which was assessed as having a failure frequency of 1.5E-03 per annum might be significantly affected, not only by time of day, but also by the age of an operator, as could the assessments of LPG transfer tasks by Munley and Bardsley (1993). Clearly, when taken together, the impact of these two EPCs, alone, could be highly significant and modify such an assessment in important respects, in ways that are not currently modelled in major hazard risk assessments.

DISCUSSION

The principal finding from the current study is that within the 25–85* years age range (*non-demented), younger operatives appear to be considerably and consistently more reliable than their elder counterparts. The finding that, for every 10 years, over this age range, human unreliability increases by about 20%, (provisional EPC multiplier of 1.2 for every ten years, age 25 to 85* years) in a compound sense, has important implications for the major hazard industries, particularly against the background of an aging workforce. Whilst it is clearly the case that experience can "counteract" the decrement in performance that is associated with age, this impact can only be relied upon to a limited extent.

As Morrow et al. (1994) have shown, expertise can "effectively" eliminate age differences in repeating air traffic control communication spoken messages as between old Pilots (mean age 66.7 years) recalling messages and young non-Pilots (mean age 28.3 years) recalling similar messages. However, based on human reliability measurements associated with simulated message read-back, it is apparent that, although "old" Pilots can be as reliable as "young" non-Pilots, there is also strong evidence that "young" persons, whether they be Pilots or non-Pilots, will always be more reliable undertaking this sort of task than the older members of their particular group.

The consequence of this particular finding and the more general finding of the overall study is that, within the 25–85* years age group (*non-demented), younger operatives do seem to be considerably and consistently more reliable than their elder counterparts, by a predictable amount.

CONCLUSIONS

This paper has shown that the impact of human performance in relation to process safety assessments can be quite considerable, and that such assessments might need to be altered in significant ways, even when quite small changes in performance are assumed. It has also shown that, once a nominal assessment has been made, this might be subject to significant variation, depending on the age of the operators and the time of day at which they perform their tasks. Further work is underway to obtain better estimates of the impact of time of day, exercise, morale and experience on human reliability, to identify additional relevant

Generic Task Types and to understand better the "additive" nature or otherwise of multiple EPCs.

Addressing human performance variability issues using these means allows them to be confronted and incorporated in process safety assessments, so that they become an integral part of the hazard management process, rather than remaining a somewhat perplexing after-thought.

REFERENCES

Ardila, A., Ostrosky-Solis, F., Rosselli, M., and Gomex, C. (2000). Age-Related Cognitive Decline during Normal Aging – The Complex Effect of Education. *Archives of Clinical Neuropsychology*, ***15***, 495–513.

Atchley, P. and Kramer, A.F. (2000). Age-related changes in the control of attention in depth. *Psychology and Aging*, ***15***, 78–87.

Baddeley, A.D., Emslie, E. and Nimmo-Smith, I. (1994). *Doors and people: a test of visual and verbal recall and recognition.* Bury St Edmunds: Thames Valley Test Company.

Bjerner, B., Holm, A., and Swensson, A. (1955). Diurnal variation of mental performance. A study of three-shift workers. *British Journal of Industrial Medicine*, ***12***, 103–110.

Blake, M.J.F. (1967). Time of day effects in performance on a range of tasks. *Psychonomic Science*, ***9***, 349–350.

Borland, R.G., Rogers, A.S., Nicholson, A.N., Pascoe, P.A. and Spencer M.B. (1986). Performance overnight in shiftworkers operating a day-night schedule. *Aviation Space and Environmental Medicine*, ***57***, 241–249.

Brookhaven National Laboratory (1989). *Risk Sensitivity to Human Error.* NUREG/CR-5319, U.S. Nuclear Regulatory Commission, Washington, D.C., 1989.

Brown, W., Crouch, D., Higgins, J., O'Hara, J. and Wong, S. (1990). *The Quantification of Human Variability and its Effect on Nuclear Power Plant Risk.* Department of Nuclear Energy, Brookhaven National Laboratory, Upton, New York, 11973, Technical Report L-1170 (1) 10/90.

Browne, R.C. (1949). The day and night performance of tele-printer switchboard operators. *Occupational Psychology*, ***23***, 121–126.

Craig, A. and Condon, R. (1985). Speed-accuracy trade-off and time of day. *Acta Psychologica*, ***58***, 115–122.

Detweiler, M.C., Hess, S.M. and Ellis, R.D. (1996). The Effects of Display Layout on Keeping Track of Visual-Spatial information. In W.A. Rogers, A.D. Fisk, and N. Walker (Eds.), Aging and Skilled Performance: Advances in theory and applications, pp. 157–184. Mahwah, NJ: Lawrence Erlbaum Associates.

Fernandes, M.A. & Moscovitch, M. (2003). Interference Effects From Divided Attention During Retrieval in Younger and Older Adults. *Psychology and Aging*, ***18***, 2, 219–230.

Folkard S. (1997). Black times: Temporal determinants of transport safety. *Accident Analysis and Prevention*, ***29***, 4, 417–430.

Gamboz, N., Russo, R., and Fox, E. (2000). Target selection difficulty, negative priming and aging. *Psychology and Aging*, ***15***, 542–550.

Glisky, E.L., Rubin, S.R., Davidson, P.S.R. (2001). Source Memory in Older Adults: An Encoding or Retrieval Problem? *Journal of Experimental Psychology*, ***27***, 5, 1131–1146.

Harris, W. (1977). Fatigue, circadian rhythm, and traffic accidents. In Mackie, R.R. (Ed.), *Vigilance: Theory, Operational performance and Physiological Correlates*, pp. 133–146, New York: Plenum Press.

Hildebrandt, G., Rohmert, W. and Rutenfranz, J. (1974). 12 and 24 hour rhythms in error frequency of locomotive drivers and the influence of tiredness. *International Journal of Chronobiology*, **2**, 175–180.

Howarth, N.L., Hefferman, C.J. and Horne, E.J. (1989). *Fatigue in Truck Accidents.* Monash University Accident Research Centre Report No. 3, February 1989.

Kausler, D.H. & Puckett, J.M. (1981). Adult age differences in memory for sex of voice. *Journal of Gerontology*, ***36***, 44–50.

Khan, F.I., Iqbal, A., Ramesh, N. and Abbasi, S.A. (2001). SCAP: a new methodology for safety management based on feedback from credible accident-probabilistic fault tree analysis system. *Journal of Hazardous Materials*, **A87**, 23–56.

Kirasic, K.C., Allen, G.L., Dobson, S.H. and Binder, K.S. (1996). Aging, Cognitive Resources and Declarative Learning. *Psychology and Aging*, ***11***, 658–670.

Kirwan, B. (1997). The validation of three Human Reliability Quantification techniques – THERP, HEART and JHEDI: Part II – Results of validation exercise. *Applied Ergonomics*, ***28***, 1, 17–25.

Kvavilashvili, L., Kornbrot, D.E., Mash, V., Cockburn, J. and Milne, A. (2005). Effects of age on prospective and retrospective memory. *Manuscript in preparation.*

McIntyre, J.S., Criak, F.I.M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology*, ***41***, 175–192.

Madden, D.J. and Langley, L.K. (2003). Age-Related Changes in Selective Attention and Perceptual Load During Visual Search. *Psychology and Aging*, ***18***, 1, 54–67.

Mantyla, T. and Backman, L. (1992). Aging and memory for expected and unexpected objects in real-world settings. *Journal of Experimental Psychology: Learning, Memory and Cognition*, ***18***, 6, 1298–1309.

Maylor, E.A. and Lavie, N. (1998). The influence of perceptual load on age differences in selective attention. *Psychology and Aging*, ***13***, 563–573.

Maylor, E.A., Vousden, J.I., Brown, G.D.A. (1999). Adult Age Differences in Short-Term memory for Serial Order: Data and a Model. *Psychology and Aging*, ***14***, 4, 572–949.

Mittler, M.M., Carskadon, M.A., Czeisler, C.A., Dement, W.C., Dinges, D.F. and Graeber, R.C. (1988). Catastrophes, Sleep, and Public Policy: Consensus Report, *Sleep*, ***11***, 100–109.

Morrow, D., Leirer, Altieri, P. and Fitzsimmons, C. (1994). When Expertise Reduces Age Differences in Performance. *Psychology and Aging*, ***9***, 1, 134–148.

Munley, G.A. and Bardsley, A.S. (1993). *A Human Factors Analysis of LPG Transfer Tasks.* SRD/HSE R599, Risley: AEATechnology. London: HMSO, ISBN 0-85356391-8.

Nicholson, A.N., Stone, B.M., Borland, R.G., and Spencer, M.B. (1984). Adaptation to irregularity of rest and activity. *Aviation, Space and Environmental Medicine*, ***55***, 102–112.

Ostrosky-Solis, F., Jaime, R.M., and Ardila, A. (1998). Memory Abilities during Normal Aging. *International Journal of Neuroscience*, ***93***, 151–162.

Parkin, A.J. and Walters, B.M. (1991). Aging, short-term memory and frontal dysfunction. *Psychobiology*, ***19***, 175–179.

Park, D.C., Smith, A.D., Dudlerm W.N. and Lafronza, V.N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, ***15***, 1185–1191.

Rankin, J.L. and Collins, M. (1985). Adult Age Differences in Memory Elaboration. *Journal of Gerontology*, ***40***, 451–458.

Roge, J., Pebayle, T., Hannachi, S.E. and Muzet, A. (2003). Effect of sleep deprivation and driving duration on the useful visual field in younger and older subjects during simulator driving. *Vision Research*, ***43***, 1465–1472.

Salthouse, T.A., Hambrick, D.Z., Lukas, K.E. and Dell, T.C. (1996). Determinants of adult age differences on synthetic work performance. *Journal of Experimental Psychology: Applied*, ***2***, 305–329.

Salthouse, T.A. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science*, ***13***, 140–144.

Samanta, P., Wong, S., Haber, S., Luckas, W., Higgins, J. and Crouch, D. (1989). *Risk Sensitivity to Human Error.* NUREG/CR-5319. Washington, D.C.: U.S. Nuclear Regulatory Commission.

Shinar, D., Tractinsky, N. and Compton, R. (2005). Effects of practice, age and tasks demands on interference from a phone while driving. *Accident Analysis and Prevention*, ***37***, 315–326.

Singer, T., Lindenbergerm U. and Baltes, P.B. (2003). Plasticity of Memory for New Learning in Very Old Age: A story of major loss? *Psychology and Aging*, ***18***, 2, 306–317.

Sliwinski, M.J., Hofer, S.M., Hall, C., Buschke, H., and Lipton, R.B. (2003). Modeling memory decling in older adults: The importance of preclinical dementia, attrition, and chronological age. *Psychology and Aging*, ***18***, 658–671.

Snaith, E.R. (1981). *The correlation between the predicted and the observed reliabilities of components, equipment and systems.* UKAEA, NCSR R18, Culcheth: National Centre of Systems Reliability.

Taylor, J. L. O'Hara, R., Mumenthaler, M.S., Rosen, A.C. and Yesavage, J.A. (2005). Cognitive Ability, Expertise, and Age Differences in Following Air Traffic Control Instructions. *Psychology and Aging*, ***10***, 1, 117–133.

Whiting, W.L. and Smith, A.D. (1997). Differential age-related processing limitations in recall and recognition tasks. *Psychology and Aging*, ***12***, 216–224.

Williams, J.C. (1985). *HEART – A Proposed Method for Assessing and Reducing Human Error.* In Proceedings of a Symposium on the Achievement of Reliability in Operating Plant, Safety and Reliability Society, 16th September 1995, Southport.

Williams, J.C. (1988). A data-based method for assessing and reducing human error to improve operational performance. In Hagen, E.W. (Ed.), *Proceedings of IEEE Conference on Human Factors in Power Plants*, pp. 436–450, Monterey, California, June 5–9.

Wong, S., Higgins, J., O'Hara, J., Crouch, D. and Luckas, W. (1990) *Risk Sensitivity to Human Error in the LaSalle PRA*, NUREG/CR-5527, U.S. Nuclear Regulatory Commission, Washington, D.C.